AD-A063 968

F/G 12/1

OPTIMAL SEQUENTIAL AND NON-SEQUENTIAL PROCEDURES FOR EVALUATING--ETC(U)

AUG 78 S GAL, C A MICCHELLI

DAAG29-75-C-0024

UNCLASSIFIED

MRC-TSR-1871

NL

1 OF 1
ADA
063968

END
DATE FILMED
3-79
DDC

(9) MRC Technical Summary Report, #1871

(6) OPTIMAL SEQUENTIAL AND NON-SEQUENTIAL PROCEDURES FOR EVALUATING A FUNCTIONAL.

(10) Shmuel Gal [redacted] Charles A. Micchelli

LEVEL II

(14) MRC-TSR

(15) DAAG29-75-C-0024

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

(11) Aug[redacted]78

(12) 24p.

(Received June 7, 1978)

Sponsored by

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

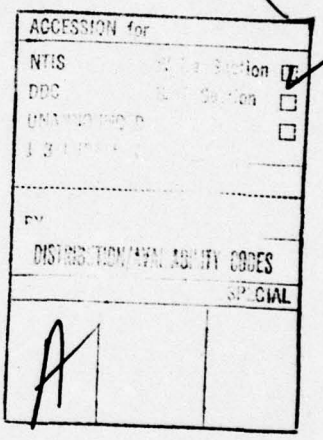
OPTIMAL SEQUENTIAL AND NON-SEQUENTIAL PROCEDURES FOR EVALUATING A FUNCTIONAL

Shmuel Gal* and Charles A. Micchelli

Technical Summary Report #1871
August 1978

ABSTRACT

In this paper we identify a class of estimation problems in which sequential estimation procedures do not yield a better rate of convergence than procedures in which all the observations are preassigned in advance.



AMS(MOS) Subject Classification - 65D99, 65J05

Key Words - Optimal estimation, sequential estimation, deterministic estimation, random estimation procedure, optimal recovery.

Work Unit Number 6 - Spline Functions and Approximation Theory

---

*The work of S. Gal was supported by IBM T. J. Watson Research Center, Yorktown Heights, New York.

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

SIGNIFICANCE AND EXPLANATION

The complexity or work required for a numerical procedure is sometimes measured in terms of the number of function evaluations required to yield an error of a given size. There are several effective algorithms which are based on a sequential method of selecting function evaluations. For instance, binary search for a root of a function or Fibonacci search for the maximum of a unimodal function are examples of sequential procedures.

In a sequential procedure the prior function values are used to pick an optimum location to evaluate the function next. An alternative to a sequential selection is to choose all the function values at once, however many have been decided to be used. Such an approach is simpler but typically yields a slower rate of convergence. For instance, the error for binary search is geometrically decreasing in the number of function values while a preassigned strategy yields only an error decreasing at a rate inversely proportional to the number of function values.

It is the purpose of this paper to identify a class of estimation problems in which sequentiality will not yield a faster rate of convergence than a deterministic choice. This class includes problems of numerical quadrature and differentiation. Thus we demonstrate that the complicated procedure of sequential estimation can be replaced by a much simpler strategy of a deterministic choice for some important problems of numerical calculation.

OPTIMAL SEQUENTIAL AND NON-SEQUENTIAL PROCEDURES FOR EVALUATING A FUNCTIONAL

Shmuel Gal* and Charles A. Micchelli

1. Framework and Defintions.

In this paper we compare certain optimal procedures for sampling a function f belonging to some prescribed class for the purpose of estimating a functional Uf. Our main objective is to identify a large collection of examples in which sequential procedures are not advantageous. Several unsettled questions whose solution would illuminate the problem studied here are outlined at the end of section 3.

We begin with a family of real valued functions defined on $[0,1]$, $F = \{f\}$, and a functional Uf defined for all $f \in F$. Assume that for any $f \in F$ we can make n observations of f at points $x_1,\ldots,x_n$ and obtain the information $f(x_1) = y_1,\ldots,f(x_n) = y_n$. The set of uncertainty in Uf is

$$Q(x;f) = \{U\varphi: \varphi \in F,\ \varphi(x_i) = f(x_i)\},\ x = (x_1,\ldots,x_n).$$

(For simplicity we shall consider the case of $x_i \in [0,1]$, but all the results hold for functions defined on any bounded subset of R^m.) As a measure of the size of this set we take

$$g(x;f) = g(x_1,\ldots,x_n;f) \tag{1}$$

$$= \sup_{\varphi \in W(x;f)} U\varphi \;-\; \inf_{\varphi \in W(x;f)} U\varphi$$

where $W(x;f) = \{\varphi:\varphi \in F,\ \varphi(x_i) = f(x_i)\}$, i.e., g is the length of the smallest interval containing $Q(x;f)$. The function $g(x;f)$ is defined for $x \in C_n = \{x = (x_1,\ldots,x_n) : 0 \le x_i \le 1\}$ and $f \in F$. We will compare three policies for choosing $x_1,\ldots,x_n$.

To this end, let E be some prescribed subset of $[0,1]$ from which we will sample a function $f \in F$. In practice, we are typically constrained to sample from some fixed finite subset of $[0,1]$. We will frequently assume E has this property.

Now, let $E_n = E \times E \times \ldots \times E$ (n times) and consider the following three procedures.

---

*The work of S. Gal was supported by IBM T. J. Watson Research Center, Yorktown Heights, New York.

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

a. Deterministic.

The accuracy guaranteed is defined as

$$d_n = \inf_{x \in E_n} \sup_{f \in F} g(x;f) \quad . \tag{2}$$

b. Random.

Random observations are preassigned by a probability distribution μ on E. The accuracy guaranteed here is

$$r_n = \inf_{\mu} \sup_{f \in F} \int_{E_n} g(x;f)\,d\mu(x) \quad . \tag{3}$$

(We will only allow probability measures $d\mu$ on E_n for which $g(x;f)$ is μ-integrable.)

c. Sequential.

A sequential search procedure is a set of n functions $h_1, h_2, \ldots, h_n$ where $h_1 = x_1$ is a constant, $x_2 = h_2(x_1, f(x_1)), \ldots, x_{i+1} = h_{i+1}(x_1, f(x_1), \ldots, x_i, f(x_i)), \ldots$, etc. This procedure produces an $h(x;f) \in E_n$, $h = (h_1, \ldots, h_n)$. The totality of sequential procedures will be denoted by S_n.

The accuracy guaranteed by a sequential procedure is given by

$$s_n = \inf_{h \in S_n} \sup_{f \in F} g(h(x;f);f) \quad . \tag{4}$$

Remark 1. Obviously $r_n \leq d_n$ and $s_n \leq d_n$ for any subset E of $[0,1]$. The following example is a simple instance in which $s_n << r_n$.

Let $E = [0,1]$, $F = \{f: f(0) = -1,\ f(1) = +1,\ f(a) \geq 0 \rightarrow f(b) > 0 \text{ for } a < b \leq 1\}$ and $Uf = \inf z$ such that $f(z) > 0$ (i.e. Uf is the "root" of f). If we let $x_0 = 0$, $x_{n+1} = 1$ then $g(x;f) = x_j - x_{j-1}$ where j is the smallest index such that $y_i = f(x_i) > 0$. Hence $\inf_x \sup_f g(x;f) = \inf_x \sup_i (x_i - x_{i-1}) = 1/n+1$. To estimate r_n we define $k_z \in F$ by $k_z(t) = -1$, $0 \leq t \leq z$, and $+1$ for $z \leq t \leq 1$. Then

$$r_n = \inf_{\mu} \sup_{f} \int_{E_n} g(x;f)d\mu(x)$$

$$\geq \inf_{\mu} \int_0^1 \int_{E_n} g(x;k_z)d\mu(x)dz$$

$$= \inf_{\mu} \int_{E_n} \sum_{i=1}^{n+1} \int_{x_{i-1}}^{x_i} (x_i - x_{i-1})dzd\mu(x)$$

$$= \inf_{\mu} \int_{E_n} \sum_{i=1}^{n+1} (x_i - x_{i-1})^2 d\mu(x) \geq 1/n+1 \quad .$$

On the other hand, the bisection (sequential) procedure yields an accuracy of $(1/2)^n$ and in fact it is easily seen that $s_n = (1/2)^n$.

In contrast to this situation we will show below that there is a wide class of problems for which $r_n \leq s_n$.

2. Linear Functionals Defined On Convex Sets.

In this section we assume that

(5) $\qquad F$ is a convex family

and

(6) $\qquad U$ is a linear functional .

Lemma 1. Suppose (5) and (6) hold. Then $g(x;f)$ is a concave function of $f \in F$.

Proof. Let $f_1, f_2 \in F$. Then for any $\epsilon > 0$ there exist $\bar{f}_i, \underline{f}_i \in F$ with $\bar{f}_i(x_j) = \underline{f}_i(x_j) = f_i(x_j)$, $i = 1, 2$, $j = 1,\ldots,n$ and

(7) $$g(x;f_i) \leq U(\bar{f}_i) - U(\underline{f}_i) + \epsilon \quad .$$

Thus by (7) and the linearity of U

$$g(x;\theta f_1 + (1-\theta)f_2) \geq U(\theta\bar{f}_1 + (1-\theta)\bar{f}_2) - U(\theta\underline{f}_1 + (1-\theta)\underline{f}_2)$$

$$= \theta[U(\bar{f}_1) - U(\underline{f}_1)] + (1-\theta)[U(\bar{f}_2) - U(\underline{f}_2)]$$

$$\geq \theta g(x;f_1) + (1-\theta)g(x;f_2) - \epsilon \quad .$$

Letting $\epsilon \to 0$ proves the lemma.

Theorem 1. If (5) and (6) hold and E is a finite subset of $[0,1]$ then

$$r_n = \sup_{f\in F} \min_{\mu} \int_{E_n} g(x;f)d\mu(x) \quad . \qquad \text{(see (3))} \tag{8}$$

This theorem is a consequence of the following version of Ky Fan's minimax theorem [2], which we state below.

Theorem A. Let S, T be sets with S a compact Hausdorff space. Let ϕ be a real-valued function defined on $S \times T$ which is continuous and convex with respect to s, and concave in t. Then there exists an $\bar{s} \in S$ such that

$$\min_{s\in S} \sup_{t\in T} \phi(s,t) = \sup_{t\in T} \phi(s,t) = \sup_{t\in T} \min_{s\in S} \phi(s,t) \quad .$$

For the proof of Theorem 1 we define

$$\phi(\mu,f) = \int_{E_n} g(x;f)d\mu(x)$$

where μ ranges over all probability measures on E_n. Given the usual topology on $\{\mu\}$ (n-dimensional simplex), the theorem follows from Lemma 1 and the definition of r_n.

This result leads us to

Corollary 1. If (5) and (6) are satisfied and E is a finite subset of $[0,1]$ then $d_n \geq s_n \geq r_n$ (Sequentiality does not help!).

Proof. Given $\epsilon > 0$ there is an $\bar{f} \in F$ with

$$\int_{E_n} g(x;\bar{f})d\mu(x) \geq r_n - \epsilon$$

for all probability measure μ. In particular, $g(x;\bar{f}) \geq r_n - \epsilon$ for all $x \in E_n$. Thus, for any sequential search procedure $h = h(x;f)$ we have $g(h(x;\bar{f});\bar{f}) \geq r_n - \epsilon$. Consequently,

$$s_n = \inf_{h\in S_n} \sup_{f\in F} g(h(x;f);f) \geq r_n - \epsilon.$$

Remark 2. A randomized sequential procedure combines, in an obvious way, the features of a random and a sequential search. Using the same argument, it follows directly that even randomized sequential procedures do not produce better accuracy than the preassigned randomized procedures considered in this paper.

We show below with an example that Theorem 1 is not valid in general for infinite subsets E of $[0,1]$.

Example 1. Let $E = [0,1]$, $F = \{f$: there is a z, $0 < z < 1$ such that f is strictly increasing in $[0,z]$ and $f(t) = f(z)$, $t \in [z,1]$ and $|f(t)| \leq 1$, $t \in [0,1]\}$ and $Uf = \lim_{t \to 1^-} f(t)$. Note that F is convex and U is linear, but E is not a finite set. Now, for $n = 2$ we have $\inf_{x_1,x_2} g(x_1,x_2;f) = 0$ and thus

$$\sup_f \inf_\mu \int g(x;f)d\mu(x) = 0 \quad .$$

However, for any μ there exists a q, $0 < q < 1$ such that $\mu(x_1$ and x_2 are not in $[q,1)) > 1/2$. Hence for ϵ small and

$$f^*(t) = \begin{cases} -1 + t\epsilon, & 0 \leq t \leq q \\ -1 + q\epsilon, & q < t < 1 \end{cases}$$

it follows that if one of the x_i's is not in $[q,1)$ then $g(x_1,x_2;f^*) \geq 2 - \epsilon$. Thus $\int g(x;f^*)d\mu \geq 1 - \epsilon/2$ and

$$\inf_\mu \sup_f \int g(x;f)d\mu \geq 1 - \epsilon/2 \quad .$$

The fact that the functions in F need not be continuous at one may seem artificial, nevertheless, it is possible to construct other examples of this type in which all the functions in F are continuous.

Remark 3. If we allow nature (our opponent) to choose an f from some fixed finite subset F' of F by means of a probability distribution $d\nu$ then the accuracy we can obtain with our best choice of $d\mu$ is

$$\min_\mu \max_\nu \int_{E_n} \int_{F'} g(x;f)d\mu(x)d\nu(f) \quad .$$

By the standard version of the minimax theorem this number equals

$$\max_\nu \min_\mu \int_{E_n} \int_{F'} g(x;f)d\mu(x)d\nu(f)$$

and optimal strategies $d\mu^*(x)$, $d\nu^*(x)$ exist even if (5) or (6) is not satisfied. In this case even though nature will choose the data according to the distribution $d\nu^*$ we know by

our previous example that a sequential procedure can guarantee a much smaller accuracy. Why is this so? The explanation to this phenomenon which is contained in Corollary 1 is as follows: When (5) and (6) are in force then nature has a universal worst function f which can be used against any sequential procedure for selecting $x = (x_1,\ldots,x_n)$. Thus we cannot hope to "learn" about $f \in F$ by using sequentiality. On the other hand, if nature has to choose among a set of functions (randomized) then we can really "learn" something about the function which was chosen. This was the case in the method of bisection.

Remark 4. Let $d\mu^*$ be the optimal probability distribution for choosing $x = (x_1,\ldots x_n)$ and suppose there is an optimal strategy of nature f^* among all $f \in F$.

Then

$$r_n = \inf_{x\in E_n} g(x,f^*) = \int_{E_n} g(x,f^*)d\mu^*(x)$$

and thus the support of μ^* is a subset of all $x \in E_n$ which satisfy $g(x;f^*) = \inf_x g(x,f^*)$. This condition may help to find the optimal search strategy, i.e., our plan against the data corresponding to the worst function.

Below we offer some further examples which show that certain extensions of Theorem 1 are not possible.

Example 2. Here we define a convex set F and a nonlinear convex functional U and demonstrate that Theorem 1 is not valid for this case.

Let $E = [0,1]$,

$$F = \{f: f \text{ concave}, |f'(x)| \leq 1, x \in [0,1]\}$$

and

$$Uf = \sup_{0\leq x\leq 1} f(x) \quad .$$

U is obviously a convex functional and F a convex set.

For $0 \leq z \leq 1$ define $\varphi_z \in F$ as $\varphi_z(x) = x$ for $0 \leq x \leq z$ and $\varphi_z(x) = 2z - x$ for $z < x \leq 1$. Thus for any n observations $x_1,\ldots,x_n$ (which we assume for convenience are ordered $0 = x_0 < x_1 <\ldots< x_n < x_{n+1} = 1$), if $x_j \leq z < x_{j+1}$ then

$g(x;\varphi_z) = z - \max(\varphi_z(x_j), \varphi_z(x_{j+1})) = \min(z - x_j, x_{j+1} - z)$.

Therefore

$$\int_0^1 g(x_1,\ldots,x_n\,;\varphi_z)\,dz$$

$$= \int_0^{x_1} (x_1-z)\,dz + \int_{x_1}^{\frac{x_1+x_2}{2}} (z-x_1)\,dz + \int_{\frac{x_1+x_2}{2}}^{x_2} (x_2-z)\,dz + \ldots + \int_{x_n}^{1} (z-x_n)\,dz$$

$$= \frac{x_1^2}{2} + \frac{(x_2-x_1)^2}{4} + \ldots + \frac{(x_n-x_{n-1})^2}{4} + \frac{(1-x_n)^2}{2}$$

$$\geq \frac{1}{4n}$$

and it follows that $r_n,\ d_n \geq \frac{1}{4n}$.

On the other hand, since all $f \in F$ are unimodal we can achieve with a Fibonacci search an accuracy of about a^n, $a \approx .62$ (the Golden section) for the location of the maximum. Moreover, since $|f'| \leq 1$ we may then locate the value of the maximum Uf, within a^n as well. Thus $s_n \leq (.62)^n << r_n,\ d_n$. A completely analogous example can be constructed with a concave functional.

Example 3. In this example $E = [0,1]$, $Uf = \int_0^1 f(x)\,dx$ and $F = \{f_z : 0 \leq z \leq 1\}$, f_z = characteristic function of $[z,1]$. Here U is a linear functional and F is a nonconvex set.

For $z \in [x_j,\ x_{j+1}]$, we have $g(x;f) = x_{j+1} - x_j$ and so

$$\int_0^1 g(x_1,\ldots,x_n;f_z)\,dz = \sum_{i=0}^{n+1} (x_i - x_{i-1})^2 \geq \frac{1}{n+1} \quad .$$

Thus $r_n \geq 1/n+1$. But, if we use the bisection sequential procedure then we may locate z with error $(1/2)^n$ and so the interval of uncertainty about Uf is $s_n = (1/2)^n << r_n$.

Examples 2 and 3 remain valid if we take E to be a finite (but large) set of points (e.g., $E = \frac{i}{m}$, $i = 0, 1,\ldots,m$).

3. Centered Sets.

We will say a set F is centered (about f_c) if there exists an $f_c \in F$ such that whenever $f \in F$ then $2f_c - f \in F$.

Lemma 2. Suppose (5) and (6) hold and F is centered about f_c. Then for all $x = (x_1,\ldots,x_n) \in E_n$ and $f \in F$

$$g(x;f) \leq g(x;f_c) \quad .$$

Proof. Since $f \in F$ if and only if $2f_c - f \in F$

$$W(2f_c - f;x) = 2f_c - W(f;x) \quad . \qquad \text{(see (1))}$$

Hence we see that $g(x;2f_c - f) = g(x;f)$ for all $f \in F$. Now, it is an easy matter, in view of Lemma 1, to prove the lemma

$$g(x;f) = \frac{1}{2}\,(g(x;f) + g(x;2f_c - f))$$

$$\leq g(x;f_c) \quad .$$

Theorem 2. If there exists an $f_c \in F$ such that for all $x \in E_n$ and $f \in F$

(9) $$g(x,f) \leq g(x,f_c)$$

then for all n

(10) $$d_n = r_n = s_n$$

and

(11) f_c is the optimal strategy of nature independently of n .

Proof: Let $v = \inf_{x \in E_n} g(x,f_c)$. It is obvious that by using f_c, nature can keep the payoff to be at least v against any randomized or sequential procedure. On the other hand for any $\varepsilon > 0$ it is possible to find an x_ε which satisfies $g(x_\varepsilon; f_c) < v + \varepsilon$, so that any $f \in F$ $g(x_\varepsilon; f) \leq g(x_\varepsilon; f_c) < v + \varepsilon$.

Thus, the deterministic search procedure x_ε keeps the payoff below $v + \varepsilon$. It follows that f_c is optimal and x_ε is ε-optimal.

Remark 5. Note that for Theorem 2 to hold, it is sufficient that condition (9) is satisfied and no other assumptions (such as linearity of U, convexity of F or finiteness of E) are necessary.

The proof of Theorem 2 actually presents a method for finding the optimal search procedure. The rule in this case is simple, we just have to find a set of points $x_1,\ldots x_n$ which minimizes the interval of uncertainty for the case in which nature uses f_c.

We also note that a related result under a stronger hypothesis appears in Bakhvalov [1].

Needless to say there are many interesting examples to which Theorem 2 applies. Below we mention one from the theory of optimal quadrature.

Example 4. Let $F = \{f: f \text{ is absolutely continuous and } |f(z) - f(y)| \le M|z - y|, 0 \le z, y \le 1\}$ and $Uf = \int_0^1 f(t)dt$. Then the assumptions of Theorem 2 hold (since 0 is the center of F) and so the observation points $x_1^*,\ldots,x_n^*$ are deterministic and preassigned and have to satisfy

$$g(x_1^*,\ldots x_n^*;0) = \min_{0\le x_1<\ldots<x_n\le 1} g(x_1,\ldots,x_n;0) \quad .$$

It easily follows that

$$g(x;0) = 2M[x_1^2 + \frac{1}{2}\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2 + (1 - x_n)^2] \quad .$$

This function has a unique minimum for $x_i^* = \frac{i-\frac{1}{2}}{n}$ $i = 1,\ldots,n$ and we conclude that

$$r_n = d_n = s_n = M/2n \quad .$$

In this example we may replace Uf by any positive linear functional $Uf = \int_0^1 f(t)d\gamma(t)$ and F by $\{f: f \text{ abs. cont. } |f'(t)| \le b(t), \text{ a.e.}\}$. Again $r_n = d_n = s_n$ but in general the optimal x_i^* are not equally spaced and difficult to find explicitly. They correspond to the minimum of

$$g(x;0) = 2\left[\int_0^{x_1}(\bar{b}(x_1)-\bar{b}(t))d\gamma(t) + \sum_{i=1}^{n-1}\left(\int_{x_i}^{z_i}(\bar{b}(t)-\bar{b}(x_i))d\gamma(t)\right.\right.$$

$$\left.\left. + \int_{z_i}^{x_{i+1}}(\bar{b}(x_{i+1}) - \bar{b}(t))d\gamma(t)\right) + \int_{x_n}^{1}(\bar{b}(t)-\bar{b}(x_n))d\gamma(t)\right]$$

where

$$\bar{b}(t) = \int_0^t b(t)dt \quad .$$

and z_i satisfies

$$\bar{b}(z_i) - \bar{b}(x_i) = \frac{1}{2}\,(\bar{b}(x_{i+1}) - \bar{b}(x_i)) \quad .$$

Our last example has the property that (5) and (6) hold and $r_n < d_n$.

In our next example we observe that the conclusion of Theorem 2 may remain valid in the absence of a center for F. This example raises the issue of the extent to which the requirement in Theorem 2 that F have a center can be weakened.

Example 5. Let $F = \{f: f \text{ increasing}, 0 \le f(x) \le 1, \text{ for } 0 \le x \le 1\}$ and $Uf = \int_0^1 f(x)dx$. Then (5) and (6) hold, however, F does not have a center. The following remarks are a formal proof of the latter statement. Assume that a center f_c exists. Since $0 \in F$ then $2f_c - 0 = 2f_c \in F$ so that $f_c \le 1/2$. But $1 \in F$ which implies $f_c \ge 1/2$. Thus $f_c = 1/2$ which is impossible because then $f(x) = x \in F$ implies $2f_c(x) - f(x) = 1 - x$ is increasing. Nevertheless we will show that $f^*(x) = x$ is a universally worst function and $d_n = r_n$.

It is easily verified that for any $x = (x_1,\ldots,x_n)$ $0 \le x_1 < \ldots < x_n \le 1$ and $f^*(x) = x$

$$g(x;f^*) = x_1^2 + (x_2-x_1)^2 + \ldots + (x_n-x_{n-1})^2 + (1-x_n)^2 \ge \frac{1}{n+1} \quad ,$$

that is, nature can guarantee a value of $\frac{1}{n+1}$ by choosing f^*. On the other hand, if we use $x_i^* = \frac{i}{n+1}$ then for any $f \in F$

$$g(x^*;f) = \frac{1}{n+1} f(\frac{1}{n+1}) + \frac{1}{n+1}\left[f(\frac{2}{n+1}) - f(\frac{1}{n+1})\right] + \ldots + \frac{1}{n+1}\left[f(\frac{n}{n+1}) - f(\frac{n-1}{n+1})\right]$$
$$+ \frac{1}{n+1}\left[1 - f(\frac{n}{n+1})\right]$$
$$= \frac{1}{n+1} \quad .$$

Thus the searcher can guarantee a value of $\frac{1}{n+1}$ by a deterministic strategy so that $d_n = r_n = s_n$. In addition, we know f^* is a universally worst function for all n. This example has been mentioned by Kiefer in [3].

Our last example has the property that (5) and (6) hold and $r_n < d_n$.

Example 6. For any interval $I \subseteq [0,1]$ we let f_I be its characteristic function. Let $I_1,\ldots,I_m$ be a partition of $[0,1]$, $I_j \cap I_k = \phi$, $j \neq k$, $\bigcup_j I_j = [0,1]$ with $I_1<\ldots<I_m$ and define

$$F = \{\sum_{i=1}^{m} y_i f_{I_i} : y_i \geq 0, \sum_{i=1}^{m} y_i = 1\}$$

(step functions). For U we choose an arbitrary linear functional given by

$$Uf = \sum_{j=1}^{n} y_j a_j$$

and as usual $E = [0,1]$. We define a mapping $i:[0,1] \to \{1,\ldots,m\}$ by the condition that $t \in I_{i(t)}$ and set $I(x) = \{i(x_1),\ldots,i(x_n)\}$, $x = (x_1,\ldots,x_n)$. Then it is easily seen that

$$g(x;f) = (1 - \sum_{j\in I(x)} y_j)(\max_{j\notin I(x)} a_j - \min_{j\notin I(x)} a_j)$$

$$= \sum_{j\notin I(x)} y_j \,(\max_{j\notin I(x)} a_j - \min_{j\notin I(x)} a_j) \quad .$$

Hence for $a_j = j$, $j = 1,\ldots,m$ we have $d_1 = m - 2$ while for the randomized procedure defined by

$$\Pr(x_1 \in I_j) = \begin{cases} \dfrac{m-1}{m^2-2m+2}, & j = 1, m \\ \dfrac{m-2}{m^2-2m+2}, & 1 < j < m \end{cases}$$

it follows that for all $f \in F$

$$\int_0^1 g(x_1;f)\,d\mu(x_1) = \frac{m-1}{m^2-2m+2}[(1-y_1) + (1-y_m)](m-2)$$

$$+ \sum_{j=2}^{m-1} \frac{m-2}{m^2-2m+2}(1-y_j)(m-1) = \frac{m^2-2m+1}{m^2-2m+2}(m-2) < m - 2 \quad .$$

Thus $r_1 < d_1$.

All our results have in effect compared d_n, r_n, s_n for a fixed n. It would be useful to determine under conditions (5) and (6) whether these quantities can be asymptotically different, in other words when is $\overline{\lim} \frac{d_n}{r_n} < \infty$.

The ideas we have presented have wider applicability. In the next section we will comment on the optimal estimation of operators from sampling in the presense of noise.

4. Extensions.

In applications it is frequently unrealistic to assume that the data $f(x_1),\ldots,f(x_n)$ is known exactly. Function values are usually only inaccurately determined as a result of either experimental or computational error. We measure these errors with a norm $\|\cdot\|$ on R^n and say that $f(x_i) = y_i + e_i$ where $\|e\| \leq 1$ (normalized), $e = (e_1,\ldots,e_n)$. In this case our uncertainty in U is the set

$$\{U\varphi: \varphi \in F,\ \varphi(x_i) = f(x_i) + e_i,\ \|e\| \leq 1\}$$

and the corresponding g is

$$g_e(x;f) = \sup_{\varphi \in H_e(x;f)} U\varphi \ - \ \inf_{\varphi \in H_e(x;f)} U\varphi$$

where

$$H_e(x;f) = \{\varphi: \varphi \in F,\ \varphi(x_i) = f(x_i) + e_i,\ \|e\| \leq 1\} \quad .$$

Continuing with this analogy we introduce d_n^e, r_n^e, s_n^e, as in Section 1, and it is easily seen that all the results of Section 2 remain valid for inaccurate data.

Our discussion in Section 2 and 3 also lends itself to the estimation of operators between linear spaces. Thus in this case U is an arbitrary mapping from a linear space X into another linear space Z. Our class F is now some subset K of X and the information about f which may be used to estimate Uf is denoted by If. Earlier If was a sample of function values $(f(x_1),\ldots,f(x_n))$. Now we allow I to be any linear mapping from X into R^n, $If = (I_1 f,\ldots,I_n f)$, $I_i: X \to R$ i.e. n linear observations of f. As before we will measure the error in the observation If with a norm $\|\cdot\|$ on R^n and thus the set of uncertainty is

$$Q(f;I) = \{U\varphi: I\varphi = If + e,\ \|e\| \leq 1,\ \varphi \in F\} \quad .$$

In this case, Q is not an interval but rather a (quite arbitrary) subset of Z. Therefore we require some measure of the size Q. For this purpose, we assume that Z is a normed linear space and recall that the Chebyshev radius of a set $T \subseteq Z$ is defined as

$$r(T) = \inf_{x\in Z} \sup_{y\in T} \|x-y\| \quad .$$

The radius of a set serves as an effective measure of its size and we let

$$g(I;f) = g(I_1,\ldots,I_n;f)$$

$$= r(Q(f;I)) \quad .$$

This definition differs by a factor of $1/2$ from the one we used earlier in (1) when $Z = R$.

An attractive feature of this way of measuring the size of Q is that it corresponds to the best way to "fill in" the diagram

$$\begin{array}{ccc} & & U \\ & X & \longrightarrow Z \\ I & \downarrow & \nearrow A \\ & Y & \end{array}$$

with a mapping A given only that $I\varphi = If + e$, $\|e\| \leq 1$ and $f \in K$. These ideas are discussed in [4] (the proofs of Lemma 1 and 2 are based on remarks in [4]).

We quote the following result from [4; p 2].

Theorem B. If U is a linear operator from X to Z and K is a convex set centered about the origin $\theta \in X$ then

$$g(I;f) \leq 2g(I;\theta) \quad .$$

With this result we may proceed, as in Section 2, to show that a sequential search cannot do better than $1/2$ a deterministic one. A sequential search in this context means that some set L of linear functionals on X is prescribed (in section 2 this was the set of point evaluations). Then a sequential method based on n observations from L is determined by a function $h = (h_1,\ldots,h_n)$, $h_i: X \times \underbrace{Y \times\ldots\times Y}_{i-1} \to L$ where $h_1 I_1 = I_1$. The function selects the information

$$h(I)f = (I_1 f,\ h_2(f,\ I_1 f),\ldots,h_n(f,\ I_1 f,\ldots,I_{n-1} f))$$

$$= (I_1 f,\ I_2 f,\ldots,I_n f)$$

and a sequential search can guarantee an error of

$$s_n = s_n(L) = \inf_h \sup_{f\in F} g(h(I)f;f)$$

while a deterministic approach yields

$$d_n = d_n(L) = \inf_{I_i \in L} \sup_{f \in F} g(I_1,\ldots,I_n;f) \quad .$$

Theorem 3. Let the hypothesis of Theorem B be satisfied. Then

$$\frac{1}{2} d_n \leq s_n \leq d_n \quad .$$

Proof. The right hand inequality is obvious. To prove the remaining inequality we let h be any sequential procedure. For $\epsilon > 0$, there exists $I^* = (I_1^*,\ldots,I_n^*)$, $I_j^* \in L$ such that

$$\inf_{I_i \in L} g(I_1,\ldots,I_n;\theta) + \epsilon \geq g(I_1^*,\ldots,I_n^*;\theta)$$

$$(\theta = \text{center of } K) \quad .$$

Thus from Theorem B

$$\sup_{f \in K} g(h(If);f)$$

$$\geq g(h(I\theta);\theta) \geq g(I_1^*,\ldots,I_n^*;\theta) - \epsilon$$

$$\geq \frac{1}{2} \sup_{f \in F} g(I_1^*,\ldots,I_n^*;f) - \epsilon$$

$$\geq \frac{1}{2} d_n - \epsilon \quad .$$

Hence, letting $\epsilon \to 0^+$ we see that any sequential procedure cannot achieve an accuracy less than $\frac{1}{2} d_n$. This proves the theorem.

When K is a unit ball given by a Hilbert space semi-norm and the data I is known exactly then (when no further assumptions on Y, Z) the factor 2 in Theorem B can be removed [5; p. 10]. For inaccurate data a similar result holds when Y can Z are Hilbert spaces [4]. In each of these cases $d_n = s_n$.

It is interesting to note that when the set L is chosen to be the set of all continuous linear functionals then d_n corresponds to the Gel'fand n-width of the set $UK = \{Uf: f \in K\}$, since

$$d_n = \inf_{I_i \in Y} g(I_1,\ldots,I_n;\theta)$$

$$= \inf_{I_i \in Y} r(Q(I_1,\ldots,I_n;\theta))$$

$$= \inf_{I_i \in Y}{}^{*} \sup_{\substack{f \in K \\ I_i f = 0}} \|Uf\|$$

$$= \inf_{I_i \in Y}{}^{*} \sup_{\substack{f \in UK \\ I_i f = 0}} \|f\|$$

the Gel'fand n-width of UK, [6].

When $K = \{x: \|x\| \leq 1\}$ is a Hilbert space unit ball and U is a bounded linear operator the Gel'fand numbers of UK are known. In particular, if U is a compact operator and

$$U^{*}U\varphi_j = \lambda_j \varphi_j, \ (\varphi_j, \varphi_k) = \delta_{jk}$$

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_n \to 0 \ .$$

Then the Gel'fand n-width equals $\lambda_{n+1}^{1/2}$ and $(\varphi_1, x),\ldots,(\varphi_n, x)$ is a best choice of elements in X^{*}.

Corollary 2. Let U be a compact linear operator from a Hilbert space X into another Hilbert space Z. Then an optimal sequential procedure for estimating Ux given $\|x\| \leq 1$ and n linear observations of x has error $\lambda_{n+1}^{1/2}$ and uses the deterministic choice $(\varphi_1, x),\ldots,(\varphi_n, x)$.

REFERENCES

1. Bakhvalov, N. S., On the optimality of linear methods for operator approximation in convex classes of functions, USSR Computational Mathematics and Mathematical Physics, 11, (1971), 244-249.
2. Fan, Ky, Minimax theorems, Proc. Nat. Acad. Sc. Vol. 39 (1953) 42-47.
3. Kiefer, J., Optimum sequential search and approximation methods under minimum regularity assumptions, J. Soc. Appl. Math., 5 (1957), 105-136.
4. Melkman, A. A. and Charles A. Micchelli, Optimal estimation of linear operators in Hilbert spaces, to appear SIAM J. Numer. Anal.
5. Micchelli, C. A. and T. J. Rivlin, A survey of optimal recovery, in Optimal Estimation in Approximation Theory, eds. C. A. Micchelli, T. J. Rivlin, Plenum Press, N.Y., 1977.
6. Singer, I., Best Approximation in Normed Linear Spaces, Springer-Verlag, New York, 1970.

SG/CAM/jvs

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 1871	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) OPTIMAL SEQUENTIAL AND NON-SEQUENTIAL PROCEDURES FOR EVALUATING A FUNCTIONAL		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Shmuel Gal and Charles A. Micchelli		8. CONTRACT OR GRANT NUMBER(s) DAAG29-75-C-0024
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 6 - Spline Functions and Approximation Theory
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE August 1978 13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Optimal estimation, sequential estimation, deterministic estimation, random estimation procedure, optimal recovery.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper we identify a class of estimation problems in which sequential estimation procedures do not yield a better rate of convergence than procedures in which all the observations are preassigned in advance.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)